# Makine Öğrenmesi

## Supervised learning, regresyon, MLE, MMSE ve MAP

Orijinal not: Ders3.pdf

**UYARI** Bu belge, el yazısı ders notlarından Claude (Anthropic) yardımıyla dijitale aktarılmıştır. Yardımcı kaynak olarak hazırlanmıştır; hata veya eksik içerebilir. Lütfen orijinal notları ve ders kaynaklarını esas alın.

**NOTICE** *This document was transcribed from handwritten lecture notes with the help of Claude (Anthropic). It is meant as a supplementary resource and may contain errors or omissions. Always check against the original notes and official course material.*

orj. s. 1

## Supervised learning:

### Sınıflandırma:

$$\mathcal{X} = \{x^t, \overset{\text{etiket}}{r^t}\}_{t=1}^{N}$$

$$x^t \in C^i \Rightarrow \text{sınıflandırma}$$

$$h_i(x^t) = \begin{cases} 1 & \text{if } x^t \in C^i \\ 0 & \text{if } x^t \in C_j,\ j \neq i \end{cases}$$

$$E\{\underset{\substack{\text{sınıflandırmadaki}\\ \text{hata}}}{\varepsilon_n} \mid \mathcal{X}\} = \sum_{t=1}^{N}\sum_{i=1}^{N} 1_{h_i(x^t)\neq r_i^t}$$

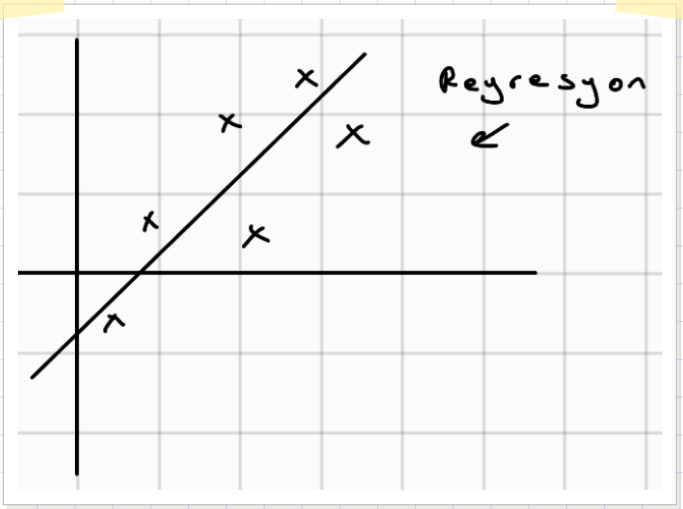
### Regresyon:

$$r^t \in \mathbb{R}$$

$$\underset{\text{kestirim}}{\hat{r}^t} = f(x^t)$$

$$r^t = \hat{r}^t = f(x^t) \Rightarrow \text{interpolasyon}$$

$$\hat{r}^t = f(x^t) + \varepsilon^{\text{görültü}} \quad \text{(regresyon)}$$




orijinal çizim, s. 1

$$\underset{\text{Expected value}}{E}\{\varepsilon_g \mid \mathcal{X}\} = \frac{1}{N}\sum_{t=1}^{N}\left[r^t - g(x^t)\right]^2$$

Loss f.

$r^t = f(x^t) + \varepsilon^{\text{görültü}}$ fonksiyonun $g(x)$ ile tahmin etmeye çalışıyoruz.

En basit modelleme: doğrusal modelleme

$$\begin{aligned} g(x) &= w_1x_1 + w_2x_2 + \cdots + w_Nx_N + w_0 \\ &= \vec{w}\cdot\vec{x} + w_0 \end{aligned}$$

orj. s. 2

$$g(x) = w_1x_1 + w_0$$

$$E\{\underset{\substack{\text{değişkenlerimiz}\\ \text{bunlar}}}{w_1, w_0} \mid \mathcal{X}\} = \frac{1}{N}\sum_{t=1}^{N}\left(r^t - (w_1x_1 + w_0)\right)^2$$

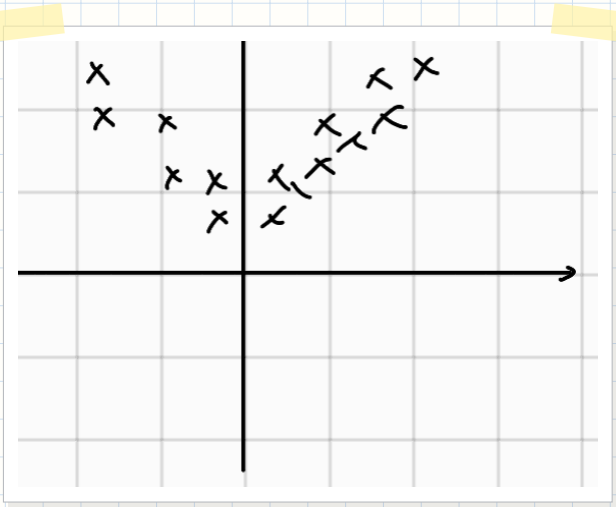
$$\frac{\partial E\{w_1, w_0 \mid \mathcal{X}\}}{\partial w_0} = -\frac{1}{N}\sum_{t=1}^{N} 2\left(r^t - (w_1x_1 + w_0)\right)$$

$$\frac{\partial E\{w_1, w_0 \mid \mathcal{X}\}}{\partial w_1} = \frac{1}{N}\sum_{t=1}^{N} 2\left[r^t - (w_1x_1 + w_0)\right](-x_1) = 0$$

$$w_1 = \frac{\sum_{t=1}^{N} x^t r^t - \bar{x}\,\bar{r}\,N}{\sum_{t=1}^{N}(x^t)^2 - N\bar{x}^2} \qquad \bar{r} = \frac{1}{N}\sum_{t=1}^{N} r^t$$

$$w_0 = \bar{r} - w_1\bar{x} \qquad \bar{x} = \frac{1}{N}\sum_{t=1}^{N} x^t$$

ya böyle olsaydı



orijinal çizim, s. 2

2, 3
1, 4
$[4,2] \rightarrow 3$
$[1,1] \rightarrow 4$

$\{x^t, r^t\}_{t=1}^N$ $g(x) = w_1 x_1 + w_0$

$$g(x) = w_2 x_1^2 + w_1 x + w_0$$

ikinci bir boyut ekledik $[x_1^2, x_1]$

orj. s. 3

## 3 adımımız var

① Model oluşturma
② Maliyet fonksiyonu

$$E[\theta \mid \mathcal{X}] = \sum_t L(r^t, g(x^t \mid \theta))$$

③ Optimizasyon Yöntemi

## Hastalık tespitinde

$$\sum_{k=1}^{K} P(C_k \mid x) = 1$$

$$R\{\alpha_i \mid x\} = \sum_{k=1}^{K} \underbrace{\lambda_{ik}}_{\text{Risk fonksiyonu}} \cdot P(C_k \mid x)$$

$C_1$ : sağlıklı
$C_2$ : kanser

| $\lambda_{ij}$ | $C_1$ | $C_2$ |
|---|---|---|
| $\alpha_1$ | 0 | 1000 |
| $\alpha_2$ | 10 | 100 |

$$\text{0/1 loss: } \lambda_{ik} = \begin{cases} 0 & \text{if } i = k \\ 1 & \text{o.w} \end{cases}$$

$$R\{\alpha_i \mid x\} = \sum_{k=1}^{K} \lambda_{ik} P(C_k \mid x) = \sum_{k \neq i} P(C_k \mid x) = 1 - P(C_i \mid x)$$

$h_i(x^t)$ — ?aidiyet fonksiyonu

$C_i$ s.t. $\arg\max_i h_i(x^t)$

orj. s. 4

çocuk bezi, tıraş bıçağı

$$\frac{P(x, y)}{P(x) \cdot P(y)}$$

$$\frac{\overset{\text{beraber alınmış}}{100}}{100 + 50 + 150} = \frac{1}{3}$$

x ile y arasındaki ilişkilerine ?bakın

## İlişki kuralları

① $P(x, y) =$

② $P(Y \mid X) = \dfrac{P(x, y)}{P(x)}$

X'leri satın ?almasının kaçında y'leri de satın ?alınmış

③ $\dfrac{P(Y \mid X)}{P(Y)} = \dfrac{P(x, y)}{P(x)\,P(y)}$

tamamen bağımsızsa $P(x, y) = P(x) \cdot P(y)$ olur

Korelasyon ile nedensellik aynı şey değil

## Bayes kuralı

$P(z \mid \underset{\text{parametre vektörü}}{\theta})$ (bunları bulmaya çalışıyoruz)

$P(x; \theta)$
θ kümesi için
X kümesinin olasılığı

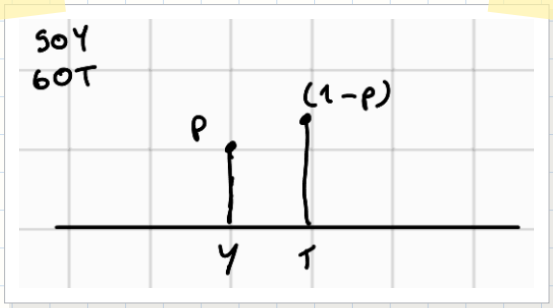
$x^t \sim P(x \mid \theta) \rightarrow$ theta parametresi ile gösterilen P dağılımına uygun

$$\theta_{MLE} = \arg\max_{\theta} P(z \mid \theta)$$

$$\ell(\theta \mid \mathcal{X}) = P(\mathcal{X} \mid \theta) = \prod_{t=1}^{N} P(x^t \mid \theta)$$

orj. s. 5

Yazı tura



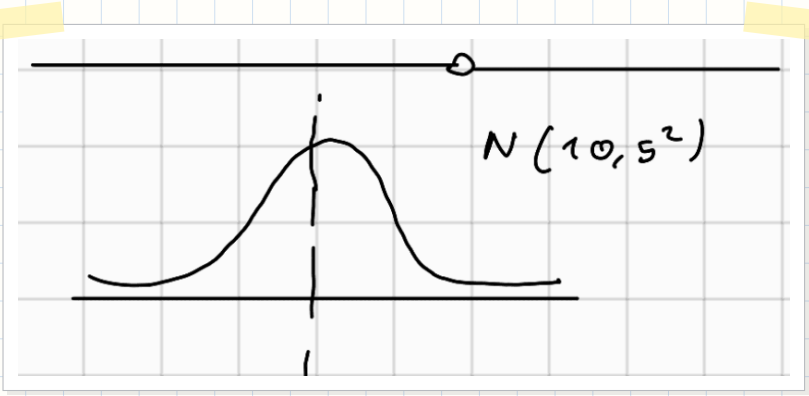

orijinal çizim, s. 5

$$\ell(\theta \mid \mathcal{X}) = P(\mathcal{X} \mid \theta) = \prod_{t=1}^{N} P(x^t \mid \theta)$$

çarpımdan kurtulmak için

$$\mathcal{L}(\theta \mid \mathcal{X}) = \log\left(\sum_{t=1}^{N} P(x^t \mid \theta)\right)^{1}$$

hem de ?üstel olasılık dağılımları çok fazla




orijinal çizim, s. 5

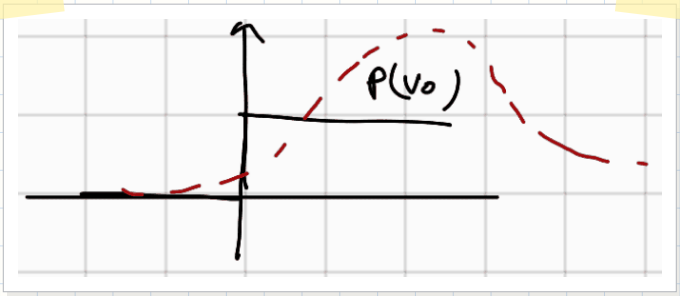
## Bayesian Theory

$$P(A \mid B) = \frac{P(B \mid A)\,P(A)}{P(B)}$$

$$\underset{\text{posterior}}{P(\mathcal{X} \mid \theta)} = \frac{\overset{\text{likelihood}}{P(\theta \mid \mathcal{X})}\ \overset{\text{prior}}{P(\mathcal{X})}}{\underset{\text{evidence}}{P(\theta)}}^{2}$$

[1] Claude notu: muhtemelen $\sum_{t=1}^{N} \log P(x^t \mid \theta)$ olmalı.

[2] Claude notu: muhtemelen $P(\theta \mid X) = \frac{P(X \mid \theta)\,P(\theta)}{P(X)}$ olmalı.




orijinal çizim, s. 5

$$\hat{\theta} = \underset{\substack{a \\ \text{karar}}}{\arg\min}\, E\{L(\theta, a) \mid X\}$$

orj. s. 6

## MMSE (Minimum mean square error).

$$C(\theta, a) = \|\theta - a\|_2^2$$

$$\bar{\theta} = E[\theta \mid \mathcal{X}]$$

$$E\{C(\theta, a) \mid \mathcal{X}\} = E\left[(\theta - a)^T(\theta - a) \mid \mathcal{X}\right]$$

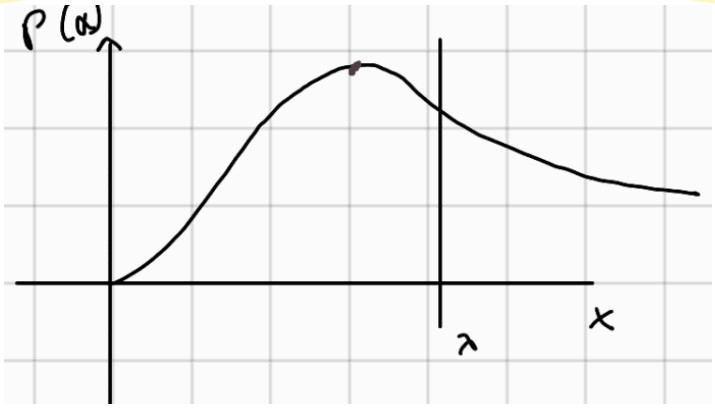
$$\frac{\partial}{\partial a} E\{(\theta - a)^T(\theta - a) \mid \mathcal{X}\} = 0$$

$$\frac{\partial}{\partial a} E\{\theta\theta^T - \theta \cdot a^T - a\theta^T + aa^T \mid \mathcal{X}\} = 0$$

$$0 - \frac{\partial\, \bar{\theta} a^T}{\partial a} - \frac{\partial\, a\bar{\theta}^T}{\partial a} + \frac{\partial (aa^T)}{\partial a} = 0$$

$$-\bar{\theta}^T - \bar{\theta} + 2a^T = 0 \Rightarrow a = \bar{\theta}$$

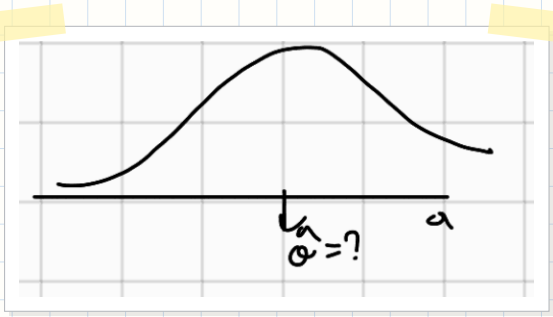
$$\hat{\theta} = \bar{\theta}$$




orijinal çizim, s. 6

# Maximum a Posteriori (Map) Estimation

$$C(\theta, a) = -\delta(\theta - a)$$




orijinal çizim, s. 6

$$E\{C(\theta, a)\} = E\{-\delta(\theta - a) \mid X\}$$
$$= -\int P(\theta \mid X)\,\delta(\theta - a)\,d\theta$$

orj. s. 7

$$= -P(\theta \mid X)\Big|_{\theta=a}$$
$$\hat{\theta}_{MAP} = \arg\min_{a} \left(-P(\theta \mid X)\Big|_{\theta=a}\right)$$
$$\hat{\theta} = \arg\max_{\theta} P(\theta \mid X)$$